Step-by-Step Guide to Installing Hadoop on Windows

An Easy Installation Process for Beginners

Author: Nikhil Shah Audience: Beginners

# Introduction

Hadoop is an open-source framework for processing and storing large datasets in a distributed computing environment. Installing Hadoop on a Windows system can be somewhat challenging, but following a clear step-by-step guide can make the process much easier.

# Prerequisites

Before starting the installation, ensure that your system meets the following prerequisites:

* Windows OS (Windows 7, 8, 10, or 11)
* Java Development Kit (JDK) installed (version 8 or later)
* At least 4GB of RAM
* Minimum 10GB of free disk space

# Step 1: Install Java Development Kit (JDK)

1.1 Download the JDK installer from the official Oracle website.

1.2 Run the installer and follow the on-screen instructions to complete the installation.

1.3 Set the JAVA\_HOME environment variable:

* Open the Control Panel and go to System and Security > System.
* Click on Advanced system settings, then click on Environment Variables.
* Under System variables, click on New.
* Enter JAVA\_HOME as the variable name and the path to your JDK installation as the variable value (e.g., C:\Program Files\Java\jdk1.8.0\_241).
* Click OK to save the variable.

# Step 2: Download Hadoop

2.1 Visit the Apache Hadoop official website.

2.2 Download the stable release of Hadoop (e.g., hadoop-3.3.1.tar.gz).

2.3 Extract the downloaded archive to a directory of your choice (e.g., C:\hadoop).

# Step 3: Configure Hadoop

3.1 Set the HADOOP\_HOME environment variable:

* Open the Control Panel and go to System and Security > System.
* Click on Advanced system settings, then click on Environment Variables.
* Under System variables, click on New.
* Enter HADOOP\_HOME as the variable name and the path to your Hadoop directory as the variable value (e.g., C:\hadoop).
* Click OK to save the variable.

3.2 Update the PATH environment variable:

* Under System variables, find the PATH variable and click on Edit.
* Add the following paths to the PATH variable (separated by semicolons):
* %HADOOP\_HOME%\bin
* %HADOOP\_HOME%\sbin

Click OK to save the updated PATH variable.

# Step 4: Setup Hadoop Configuration Files

4.1 Navigate to the Hadoop configuration directory (e.g., C:\hadoop\etc\hadoop).

4.2 Edit the following configuration files:

## core-site.xml

Add the following configuration inside the tags:

fs.defaultFS

hdfs://localhost:9000

## hdfs-site.xml

Add the following configuration inside the tags:

dfs.replication

1

dfs.namenode.name.dir

file:///C:/hadoop/data/namenode

dfs.datanode.data.dir

file:///C:/hadoop/data/datanode

## mapred-site.xml

Rename mapred-site.xml.template to mapred-site.xml and add the following configuration inside the tags:

mapreduce.framework.name

yarn

## yarn-site.xml

Add the following configuration inside the tags:

yarn.nodemanager.aux-services

mapreduce\_shuffle

# Step 5: Format the Namenode

5.1 Open a command prompt as an administrator.

5.2 Run the following command to format the namenode:

hdfs namenode -format

# Step 6: Start Hadoop Services

6.1 Start the HDFS services by running the following commands:

start-dfs.cmd

6.2 Start the YARN services by running the following command:

start-yarn.cmd

# Step 7: Access Hadoop

7.1 Open your web browser and navigate to the following URLs to verify that Hadoop services are running correctly:

* NameNode: [URL]/
* ResourceManager: [URL]/

# Conclusion

By following these steps, you have successfully installed Hadoop on your Windows system. You can now begin exploring the capabilities of Hadoop and its ecosystem to process and analyze large datasets. Happy Hadooping!